HA-SAM: Hierarchically Adapting SAM for Nerve Segmentation in Ultrasound Images

Zihao Peng[1], Susu Kang[1], Xuping Huang[1], Xucheng Xiang[1], Gengyu He[1], Tianzhu Liu[2], Wei Mei[2(✉)] and Shan Tan[1(✉)]

[1] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China
{pengzh42,kss2020,d202481550,hust_xcxiang,m202373451,shantan}@hust.edu.cn

[2] Department of Anaesthesiology Tongji Hospital, Tongji Medical College Huazhong University of Science and Technology, Jiefang Ave. 1095#, Wuhan, China
liutzh@tjh.tjmu.edu.cn,wmei@hust.edu.cn

Abstract. Automatic ultrasound nerve localization algorithm is crucial in nerve block procedures and neuropathy detection. However, the performance of existing approaches is typically constrained by the limited scale of ultrasound image datasets. While adapting from large scale models such as Segment Anything Model (SAM) has demonstrated remarkable performance on medical images, its effectiveness heavily relies on extensive datasets and substantial computational resources. This presents significant challenges for adapting SAM to ultrasound image segmentation. To address these challenges, we propose a novel parameter- and data-efficient adaptation method called Hierarchical Adapter. Specifically, the Hierarchical Adapter can flexibly adjust the number of fine-tuning parameters to optimize the exploitation of data and computational resources. In addition, we observe the depth-dependent difficulty for adapting different Transformer blocks of SAM. Therefore, we insert Hierarchical Adapters with varying sizes into transformer layers at different depths of the SAM encoder, optimizing the distribution of trainable parameters. This design significantly improves the parameter-efficiency during adaptation while simultaneously enhancing segmentation performance. Compared to state-of-the-art methods, our model reduces training parameter requirements by more than half while still achieving an approximately 1.5% improvement in Dice score on two ultrasound nerve datasets.

Keywords: Large Scale Model · Segment Anything Model · Ultrasound Nerve Image Segment.

1 Introduction

Ultrasound is widely used in medical imaging for diagnosis and surgical procedures due to its non-invasive nature, real-time imaging capability, and low cost.

Identifying and segmenting nerve structures is important for safe anesthesia and surgery [22]. It is also crucial in creating personal treatment plans for neurological rehabilitation [2]. However, the limited dataset scale poses a crucial challenge in training an automatic segmentation algorithm for accurate ultrasound nerve segmentation [27].

Recent studies have demonstrated that adapting large-scale pre-trained models can significantly improve the performance on downstream tasks [7]. With the advancement of large-scale models, the Segment Anything Model (SAM) has garnered significant attention in image segmentation. However, SAM's training set (SA-1B) contains a very few ultrasound images, resulting in suboptimal performance for ultrasound segmentation [23]. Despite SAM's strong generalization ability, adapting its millions of parameters [19, 21, 30] is computationally expensive and typically requires large-scale datasets, which is particularly challenging given the limited availability of nerve ultrasound data.

To overcome these challenges, we propose a highly data-efficient and computationally efficient method to effectively adapt SAM for nerve ultrasound image segmentation. Particularly, we propose a novel Hierarchical Adapter method and Hierarchical Adapter SAM (HA-SAM) to improve the efficiency of adapter parameters in Parameter-Efficient Fine-Tuning (PEFT). Our analysis reveals that different Transformer layers in SAM's image encoder require varying degrees of adaptation, with deeper layers benefiting from more parameters. To optimize adaptation, we assign smaller adapters to shallow layers and larger adapters to deeper layers, enhancing SAM's segmentation performance on ultrasound images without requiring additional training data or computational resources. Experimental results confirm the superior efficiency and effectiveness of our method for ultrasound image segmentation. We also collect and label the NBSBP dataset, which is helped by doctors at TongJi Hospital and is the first supraclavicular brachial plexus dataset for nerve block segmentation, called Neural Blockade Supraclavicular Brachial Plexus (NBSBP).

In summary, our key contributions are as follows: We introduced HA-SAM, a resource-efficient method specifically designed for small ultrasound neural datasets, significantly reducing computational and data-scale demands while improving segmentation performance.

We validated the effectiveness of HA-SAM through extensive experiments on ultrasound segmentation datasets, demonstrating its clear advantages over both task-specific and SAM-based methods.

2 Related Work

SAM in Medical Image Segmentation. Meta AI developed the SAM, which is considered a milestone in visual foundational models [15]. But, even though SAM works well on natural images, it does not work well for medical image segmentation [23]. Many studies have attempted to transfer SAM to the field of medical imaging. For example, Med-SA [30] adds medical image details using Space-Depth Transpose and the Hyper-Prompting Adapter. SAMed [31] uses the

LoRA method to help SAM perform semantic segmentation on medical images. MedSAM [21] collects a large set of medical images and fine-tunes parts of the model on this data. These methods do not work well for ultrasound as they do not target ultrasound images. SFRecSAM [32] adds frequency features for breast ultrasound images, and SAMUS [19] adds local features for ultrasound images. However, they need many training resources. Our HA-SAM method in this paper reduces the training resources needed.

Parameter Efficient Fine-Tuning. PEFT fine-tunes only a few parameters in large models and keeps most of them fixed. This method reduces the need for computation and storage. [9]. Adaptation methods [10, 29] add adapter modules into the transformer layers and fine-tune just the modules. AdaptFormer [5] uses this idea in Vision Transformer. LoRA [12] decomposes certain weight matrices in a pre-trained model into two smaller matrices, updating only these during training. Prefix tuning [18] and prompt tuning [17] add trainable tokens to the input or to the middle of the sequence. Visual Prompt Tuning [14] uses prompt tuning for image classification.

3 Method

3.1 The Overall Architecture of HA-SAM

Our method is inspired by migrating large models to downstream tasks. Layered fine-tuning, a technique that adjusts the unfreezing strategy on a layer-by-layer basis, enables flexible adaptation to tasks that differ significantly from pre-training data, aligning with our goal of adapting SAM to ultrasound nerve images [25]. Instead of applying uniform fine-tuning, this approach updates each layer to varying degrees based on model hierarchy. Typically, high-level layers near the output are unfrozen for task-specific adaptation, while low-level layers remain frozen to retain general features [11, 28]. Layered fine-tuning has proven effective in vision tasks, particularly under distribution shifts [6, 16, 20]. Building on this, we propose the Hierarchical Adapter, which introduces adapter blocks of different sizes in both high- and low-level layers, enabling fine-tuning at varying depths.

As depicted in Fig. 1 , HA-SAM retains SAM's overall architecture, including the prompt encoder and mask decoder, both of which remain frozen during training. The ultrasound neural image and a point prompt are input into the image encoder to extract features, which are then combined with prompt features and fed into the mask decoder to generate a segmentation result.

The image encoder is enhanced by integrating the Hierarchical Adapter into its Transformer blocks, improving upon the adaptation method. Small adapters are added to shallow blocks, while larger adapters are used in deep blocks to efficiently adapt high-level features while making minimal changes to low-level features. This approach leverages SAM's generalization ability, introduces fewer parameters, optimizes training on a small ultrasound dataset, and reduces resource consumption.

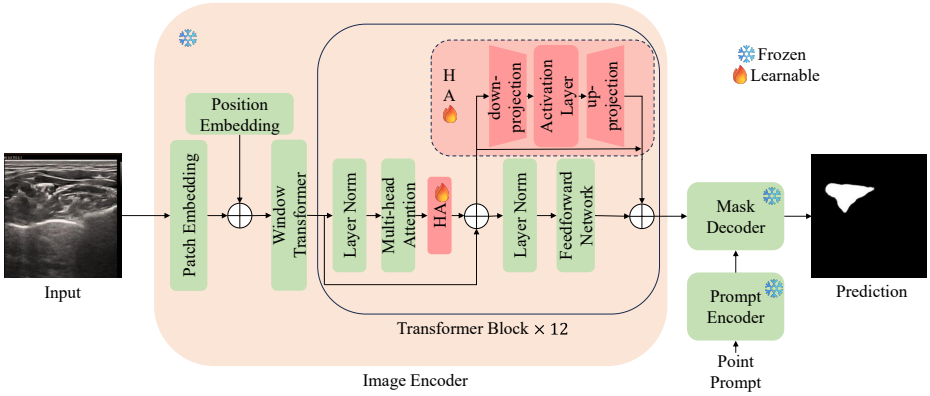

Fig. 1. Overview of HA-SAM.

3.2 Hierarchical Adapter

The structure of Hierarchical Adapter is the bottleneck structure shown in Fig. 1. The purpose of this is to limit the number of parameters introduced. First, a down-projection $W_{down} \in \mathbb{R}^{d \times r}$ is used to project the input d-dimensional original features into a low-dimensional space specified by the bottleneck coefficient r, followed by a nonlinear activation function $f(\cdot)$, and finally an up-projection $W_{up} \in \mathbb{R}^{r \times d}$ to restore the output to the original d-dimensional original features. By setting $r \ll d$, we can limit the amount of added parameters and balance parameter efficiency with performance. At the same time, we set residual connections within each hierarchical adapter to ensure the stability of feature extraction. We integrate hierarchical adapters into the transformer layers. The first adapter is positioned after the multi-head attention in the attention layer, connected in series with the multi-head self-attention module, and precedes the residual connection; the second is placed within the residual connection of the feed forward layer, running in parallel with the feed forward module.

To more efficiently utilize the parameters added to the image encoder, we drew inspiration from the concept of layered fine-tuning. For Transformer layers at different depths, we adjust the parameter size of the inserted hierarchical adapters by tuning the bottleneck coefficient r. Specifically, as illustrated in Fig. 2, we reduced the size of the adapter used to fine-tune the shallow transformer blocks, while increasing the size of the adapter used to fine-tune the deep transformer blocks. Our current hierarchical strategy follows an arithmetic progression, where the size of the Hierarchical Adapter increases arithmetically with the depth of the transformer layer. Let i represent the number of layers in the Hierarchical Adapter. For the shallow Hierarchical Adapter ($1 \leq i \leq 6$), its bottleneck coefficient r_i can be expressed as:

$$r_i = d \times r \times (\frac{i}{6}). \tag{1}$$

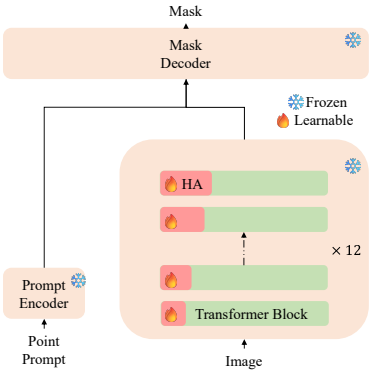

Fig. 2. Hierarchical Adapter. The size of HA represents the learnable parameter capacity.

For the deep Hierarchical Adapter ($7 \leq i \leq 12$), the bottleneck coefficient r_i can be expressed as:

$$r_i = d \times r \times (\frac{i-1}{6}), \tag{2}$$

where d is the dimension of the network itself, and r is the standard bottleneck coefficient(with $r = r_6 = r_7$). In this approach, while keeping the total number of trainable parameters constant, we allocate more parameters to adjust high-level features and reduce parameter allocation for low-level features. This strategy introduces ultrasound image-specific information into SAM, enhancing its segmentation performance on ultrasound images while preserving its strong generalization capabilities. Thus, if $h_0^{'}$ represents the embedding from the first Transformer block, the final output embedding h_{12} can be expressed as:

$$\begin{aligned} h_{12} = h_{11}^{'} + f(h_{10}^{'} + f(\cdots(h_0^{'} + f(h_0^{'} W_{down}^{1}) W_{up}^{1}) \\ \cdots W_{down}^{11}) W_{up}^{11}) W_{down}^{12}) W_{up}^{12}), \end{aligned} \tag{3}$$

where $h_i^{'}$ represents the input to the i-th layer, h_i represents the output after passing through the i-th layer Transformer Block, which can be represented as:

$$h_i = h_{i-1}^{'} + f(h_{i-1}^{'} W_{down}^{i}) W_{up}^{i}. \tag{4}$$

The above formula (3) expands h_i using formula (4) in the activation function. W_{down}^{i} represents the down-projection of the i-th layer Hierarchical Adapter, W_{up}^{i} represents the up-projection of the i-th layer Hierarchical Adapter, while $f(\cdot)$ denotes the nonlinear activation function, for which we use ReLU.

3.3 Prompt Encoder

In our study, we use point prompts to simulate how doctors interact with areas during segmentation. To avoid directly using the ground truth for click prompts,

we develop a simple target localization network to predict point coordinates. We employ a lightweight ResNet-18, by modifying its fully connected layer to identify the target area and output point coordinates. During training, we set the ground truth center as the target and optimized the network using Mean Squared Error (MSE) Loss. During testing, a prediction is considered correct if it falls within the foreground of the ground truth. Our results show that the network achieves an accuracy of 99.97%, meeting doctors' requirements for click-based prompts in surgical settings.

4 Experiments

4.1 Datasets and Implementation Details

In our experiments, we evaluate segmentation performance on two ultrasound nerve datasets and an additional breast cancer ultrasound dataset. The first is the neural blockade supraclavicular brachial plexus dataset consisting of 1,500 images from 50 patients, specifically designed for nerve block segmentation (NB-SBP). The second is the Kaggle BP Dataset [24], a publicly available dataset containing 2,500 images for brachial plexus segmentation. Additionally, we tested our method on the BUSI dataset [1], which includes nearly 650 ultrasound images of breast cancer from women aged 25 to 75 years. This diverse dataset selection ensures a comprehensive evaluation of our method across various ultrasound segmentation tasks.

We split each dataset randomly into training, validation, and test sets with a 7 : 1.5 : 1.5 ratio. We compare our method with the following ten SOTA methods: SAM [15], Med-SA [30], SAMed [31], MedSAM [21], SAMUS [19], Unet [26], SwinUnet [3], TransUnet [4], H2Former [8], and MissFormer [13].

We train the model on a single NVIDIA 2080TI GPU, using the ViT-B version of SAM. During training, only the hierarchical adapter's parameters can be learned. To match SAM's image size limit, we adjust the resolution of ultrasound images to 1024×1024. We select the Adam optimizer for training and set the initial learning rate to 0.0005, batch size to 1, and epoch to 400. The evaluation indicators are dice, IoU and HD95, all of which are averaged. Additionally, we perform a two-sample t-test to calculate the p-value, assessing the statistical significance of the observed differences in model performance.

4.2 Compare HA-SAM with SOTA Models

Table 1 summarizes the performance of our method and other SOTA models on three datasets. HA-SAM outperforms existing methods, with particularly strong results on NBSBP. Compared to vanilla SAM, it improves the Dice score by 30%, surpasses task-specific methods by about 5%, and outperforms other SAM-based SOTA models by an average of 1%. We also report the number of trainable parameters in Table 1. HA-SAM has the fewest among all compared methods, with only 6.82M. This superior performance highlights the effectiveness of the Hierarchical Adapter, which efficiently integrates ultrasound-specific

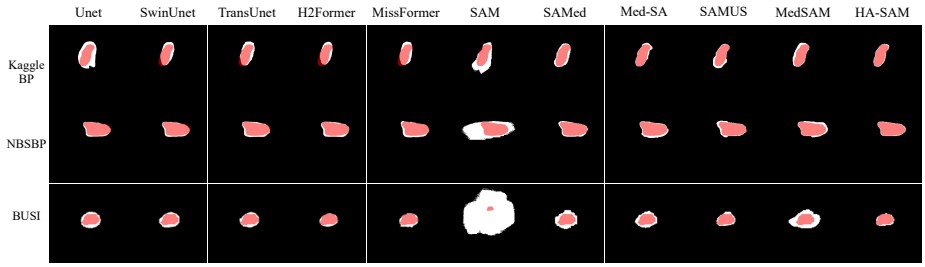

Fig. 3. Qualitative results of HA-SAM and SOTA models. For clarity in comparisons, the ground truth is highlighted in pink.

information while using a layer-wise adapter structure to maximize parameter efficiency. As a result, HA-SAM reduces computational cost while improving segmentation performance on data-limited ultrasound nerve datasets. Additionally, its parameter-efficient design makes it highly practical in low-resource environments. To assess generalization, we tested HA-SAM on the BUSI dataset, confirming its effectiveness beyond nerve segmentation. Fig. 3 presents qualitative results, showing improved segmentation quality.

Table 1. Comparison of Different Models. An asterisk (*) after the data indicates a statistically significant difference at the 5% significance level ($p<$ 0.05).

Model	Train Para.	NBSBP			KaggleBP			BUSI		
		Dice	IoU	HD95	Dice	IoU	HD95	Dice	IoU	HD95
Unet	23.53	70.20*	55.16*	16.15*	69.63*	58.92*	23.43*	71.24*	56.36*	12.01*
SwinUnet	27.17	69.58*	54.52*	17.18*	76.40*	64.76*	20.65*	73.88*	58.49*	15.94*
TransUnet	105.28	76.45*	62.72*	14.22	77.36*	65.54*	17.98*	75.62*	65.73*	**4.86**
H2Former	33.87	79.05*	66.34*	16.67*	76.89*	65.28*	27.54*	74.79*	64.97*	4.90
MissFormer	42.46	78.45*	65.67*	**12.98**	76.46*	64.98*	22.79*	75.23*	65.10*	6.75
SAM	0.00	52.20*	43.06*	150.46*	46.76*	38.53*	61.18*	40.79*	33.73*	294.95*
SAMed	18.81	81.36*	68.13*	26.92*	78.80*	67.32*	41.43*	76.74*	62.76*	14.44*
Med-SA	13.00	81.84*	68.59*	16.34	78.12*	66.39*	15.37*	77.02*	65.37*	7.88
SAMUS	130.10	82.53*	70.02*	18.27*	78.31*	66.47*	**12.75**	77.81	66.18	11.71*
MedSAM	93.73	80.73*	67.72*	17.56*	75.79*	64.18*	14.63*	75.96*	61.80*	19.87*
HA-SAM	**6.82**	**83.04**	**70.43**	15.08	**79.44**	**66.73**	13.06	**78.56**	**65.67**	9.62

4.3 Changes in Feature Distribution Across Different Layers

As shown in Fig. 4, we compare the output feature value distributions from shallow and deep Transformer blocks, both with and without the hierarchical strategy. The results show that the hierarchical strategy has little effect on shallow

features, but it significantly alters the feature distribution in deep Transformer blocks. These results highlight the difference in migration difficulty between low-level and high-level features. Our strategy improves the efficiency of high-level feature migration while having minimal impact on low-level features. Additionally, this improvement is achieved without adding extra training parameters, demonstrating the effectiveness of the hierarchical design in the adapter.

4.4 Ablation Study

Table 2 summarizes the ablation experiments, primarily evaluated using the Dice metric. We test several configurations, including reversing the hierarchical adapter size (using larger adapters in shallow layers and smaller ones in deeper layers), removing the first and last hierarchical adapters, and applying different hierarchical strategies, which change the design of bottleneck coefficient r_i. Strategy 1 follows $r_i = 2^{\frac{i-6}{6}}$, and Strategy 2 follows $r_i = e^{\frac{i-6}{6}}$. The results show that Transformer blocks at different depths have varying migration challenges and our hierarchical adapter method improves performance by prioritizing high-level feature migration. The different hierarchical strategies have a minimal impact on performance, as they introduce only slight changes in the number of parameters.

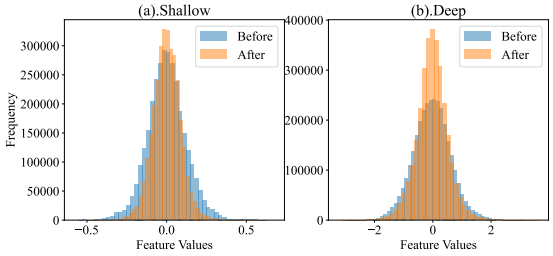

Fig. 4. Histograms of eigenvalues before and after stratification in shallow and deep layers.

Table 2. Ablation Study.

Method	Dice
Reversing HA sizes	82.55
Removing the first layer	80.75
Removing the last layer	80.25
Strategy 1	82.74
Strategy 2	82.89
HA-SAM	**83.04**

5 Conclusion

This study introduces HA-SAM, an adaptation of SAM for ultrasound nerve segmentation. By adjusting adapter parameters at different depths, HA-SAM enhances segmentation performance without increasing the number of trainable parameters. We evaluate its effectiveness on both private and public ultrasound nerve datasets and further test its generalization ability on a public ultrasound breast cancer dataset. Experimental results demonstrate that HA-SAM effectively adapts SAM to small datasets, achieving superior performance in resource-constrained environments with limited data.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. Bowness, J.S., Burckett-St Laurent, D., Hernandez, N., Keane, P.A., Lobo, C., Margetts, S., Moka, E., Pawa, A., Rosenblatt, M., Sleep, N., et al.: Assistive artificial intelligence for ultrasound image interpretation in regional anaesthesia: an external validation study. British journal of anaesthesia **130**(2), 217–225 (2023)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
4. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
5. Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., Luo, P.: Adaptformer: Adapting vision transformers for scalable visual recognition. Advances in Neural Information Processing Systems **35**, 16664–16678 (2022)
6. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: International conference on machine learning. pp. 647–655. PMLR (2014)
7. Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., et al.: Pre-trained models: Past, present and future. AI Open **2**, 225–250 (2021)
8. He, A., Wang, K., Li, T., Du, C., Xia, S., Fu, H.: H2former: An efficient hierarchical hybrid transformer for medical image segmentation. IEEE Transactions on Medical Imaging **42**(9), 2763–2775 (2023)
9. He, J., Zhou, C., Ma, X., Berg-Kirkpatrick, T., Neubig, G.: Towards a unified view of parameter-efficient transfer learning. arXiv preprint arXiv:2110.04366 (2021)
10. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
11. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146 (2018)
12. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
13. Huang, X., Deng, Z., Li, D., Yuan, X.: Missformer: An effective medical image segmentation transformer. arXiv preprint arXiv:2109.07162 (2021)
14. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision. pp. 709–727. Springer (2022)
15. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)

16. Lee, Y., Chen, A.S., Tajwar, F., Kumar, A., Yao, H., Liang, P., Finn, C.: Surgical fine-tuning improves adaptation to distribution shifts. arXiv preprint arXiv:2210.11466 (2022)
17. Lester, B., Al-Rfou, R., Constant, N.: The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691 (2021)
18. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. arXiv preprint arXiv:2101.00190 (2021)
19. Lin, X., Xiang, Y., Zhang, L., Yang, X., Yan, Z., Yu, L.: Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation. arXiv preprint arXiv:2309.06824 (2023)
20. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
21. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
22. Marhofer, P., Greher, M., Kapral, S.: Ultrasound guidance in regional anaesthesia. British journal of anaesthesia **94**(1), 7–17 (2005)
23. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: an experimental study. Medical Image Analysis **89**, 102918 (2023)
24. Montoya, A., Hasnin, kaggle446, shirzad, Cukierski, W., yffud: Ultrasound nerve segmentation. https://kaggle.com/competitions/ultrasound-nerve-segmentation (2016)
25. Ro, Y., Choi, J.Y.: Autolr: Layer-wise pruning and auto-tuning of learning rates in fine-tuning of deep networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 2486–2494 (2021)
26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
27. Song, K., Feng, J., Chen, D.: A survey on deep learning in medical ultrasound imaging. Frontiers in Physics **12**, 1398393 (2024)
28. Sun, Y., Xie, Y., Ding, B., Li, Y., Zhang, J.: Exploring selective layer fine-tuning in federated learning. arXiv preprint arXiv:2408.15600 (2024)
29. Sung, Y.L., Cho, J., Bansal, M.: Lst: Ladder side-tuning for parameter and memory efficient transfer learning. Advances in Neural Information Processing Systems **35**, 12991–13005 (2022)
30. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
31. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
32. Zhang, W., Wu, H., Qin, J.: Domesticating sam for breast ultrasound image segmentation via spatial-frequency fusion and uncertainty correction. In: European Conference on Computer Vision. pp. 20–37. Springer (2024)